# コンジョイント分析における汎用的なモデル選択基準及び高次元データのモデル推定

代表研究者　　黒　沢　健　　　東京理科大学　理学部　准教授

## 1 はじめに

近年、ビックデータの活用に伴い、情報通信サービスに関するユーザの嗜好をモデリングし、サービスレコメンデーションなど、ユーザの意識の裏で統計モデルが活用されている。また、ユーザから集めたコンジョイント分析を可能とするような市場調査データに基づき、情報通信サービスのユーザの行動データ（Revealed Preference data=RP データ）もしくは嗜好データ（Stated Preference SP データ）から統計モデルを構築し、サービスの向上やサービスの需要予測がおこなわれている。このように、情報通信サービスに関するユーザの行動モデリングやその他の関連した各種の統計モデルは、ユーザの利便性の向上という点だけではなく、情報通信サービスを提供するプロバイダーからの視点においても重要な意味をなしている。

特にコンジョイント分析では、ロジットモデルもしくは離散選択モデルと呼ばれるモデルによってユーザの嗜好を表現するモデリングがおこなわれる。例えば、ADSL から光ファイバの移行期において、依田[1]によってブロードバンド・エコノミクスと呼ばれる分野で情報通信サービスの評価が行われている。本著者も共著者と共に、IP 電話やブロードバンドサービスのユーザ評価及び需要予測の観点からコンジョイント分析を用いたモデル評価を多数実施している[2-7]。そこで、問題となるのが統計モデルの妥当性の判断である。実用上、統計モデルの妥当性の判断においては、昔から使われているある基準が存在するが、その値の妥当性については理論的な背景の欠如から疑問視する声も存在する。そこで本検討では、1 つ目の課題として、コンジョイント分析に関連するモデル評価基準の検討を進める。実例として、スパムメールの判定モデルに関する評価も 3-2 節で紹介する。

また、もう一つの問題として、情報通信データの巨大化に伴って発生するデータの高次元化の問題である。センサーデータや IoT デバイスの台頭に伴い、情報通信機器が持つデータが巨大化していることは誰も否定できない。こうしたデータの活用に際し、データの高次元化はデータ分析者にとっての問題となり得る。データの高次元化に伴い、データ自体が従属的な関係を持ち、有効なデータを活用しにくい現状にある。近年、AI や深層ラーニングなどの基幹技術として位置づけられる機械学習の分野で Lasso と呼ばれる縮小推定法が脚光を浴びている。統計モデルの表現において、データの高次元化に伴って推定するべき統計モデルのパラメータ空間が高次元化して、その推定を Lasso 等の縮小推定法に頼ることがある。本検討では、Lasso とは異なる Liu 推定と呼ばれる縮小推定法に関する新しい推定法を検討する。

## 2 関連研究

1 節で述べたように 2 つの課題に対して 2-1 節、2-2 節の各節で既存の手法の問題点を簡単に紹介する。

### 2-1 McFadden 決定係数

１節で述べたように、依田[1]や著者およびその共著者[2-7]で用いられている統計モデルは、主にロジットモデルと呼ばれるロジスティック回帰を多値化したものであり、その一般的な階層化ロジットモデルやその他の様々な一般化されたモデルが利用されている。モデルは突然一つに定まるわけではなく、様々な仮定の下で複数の統計モデルを構築するのが一般的な分析の流れである。複数の統計モデルを用意する理由としては、ある原因が結果に対して効果的かどうかを判断するために、その変数を含む統計モデルとその変数を含まない統計モデルを用意し、両方の統計モデルを比較する必要がある。実際には原因（変数）は一つではないため、複数の変数の組み合わせから発生する多くの統計モデルが仮定される。そして最終的にそれらの中から最も良いモデルを**選択**するのである。統計モデルの選択は、様々な方法によっておこなうことができるが、代表的なものは AIC[8]と呼ばれる値である。これは分野を問わず、データが発生する確率モデルを仮定して作られる尤度関数を用いた統計量であり、この値を比べることで仮定した複数の統計モデルの中から尤も良い統計モデルを**選択**する。AIC もしくはそこから派生された統計量は有用であり、その利便性の良さについては議論の余地がない。しかし、この AIC の目的は選択であって、**モデルの評価ではない**ことに注意

が必要である。なぜなら、AIC は相対基準であって、値そのものには意味がないからである。つまり、複数の統計モデルから計算されたAIC を比べ、もっと値が低い統計モデルを最も良いモデルとして判断するのがAIC の使い方であって、実際にそのモデルが良かったのかを判断するものではない。つまり、そもそも比べているモデルがドングリの背比べであって、どれもモデルとしての当てはまりが十分ではないものを比べても意味がない場合がある。そしてその選択によって得られたモデルから考察された結論は意味をなさないばかりか、結果を婉曲してしまう可能性すらあり得る。モデル自体が良いモデルなのか判断するモデル評価基準も統計モデルの決定において大事な役割を果たす。よって、本検討の対象はモデル選択ではなく、**モデル評価**であることに注意する。ロジスティック回帰のようなモデル評価基準として、判別率・的中率などの指標もあるが、頑健性という観点から的中率などの指標は使われることは多くない。特にロジスティック回帰モデルやロジットモデル(実際にはこれらのモデルに限られたモデル評価ではないが)に共通的なモデル評価基準とは、離散選択モデルに関連する研究でノーベル経済学賞を受賞したMcFadden によって提案されたMcFadden 決定係数[9]と呼ばれるモデル評価尺度である。これは線形モデルで使われている決定係数と同様に、０から１の間の値をとる。この評価尺度は 0 から 1 の値をとる統計量であり、値が大きいほどモデルの良さを表すため、モデル評価基準として用いられてきている。しかし、この McFadden 決定係数は尤度関数を用いて擬似的に決定係数を模擬しただけで、その値に理論的な裏付けは存在しない。つまり、値自体に解釈を与えることが難しいのである。この問題点を改善するため、線形モデルを仮定したときにそのモデル評価基準が決定係数と一致するように改善したCox and Snell [10]によるモデル評価基準が存在する。しかし、これにも問題があり、この評価基準には明示的な上限が存在し、モデルが完全にデータに適合していても選択したモデルによっては値が 1 をとらないという問題があった。それを調整するために、Nagelkerke [11]によって上限が１になるように調整が行われたが、Nagelkerke によるモデル評価基準は、線形モデルを仮定したときに再び決定係数とは一致せず、Cox and Snell によって提案されたモデル評価基準のアピールポイントを消してしまった。このような状況ではあるが、3 つの尺度は依然としてロジットモデルのモデル評価尺度で使われている現状にあり、それぞれの値は違う値を返すが、その解釈を与えることができず、結局、統計モデルが良かったのか悪かったのかも判断するのは難しく、その尺度によって統計モデルが与えられ、それが正しいモデルと信じて応用されているのが現状である。そこで、ロジスティック回帰モデルを含むような一般化線形モデル　[12]のクラスに対応したモデル評価基準 RCC が Zheng and Agresti[13]によって提案された。更には Eshima and Tahata によって ECC, ECD といった様々なモデル選択基準が提案[14,15,16]された。本検討はこれらの新しい基準に関する検討とその応用に関する検討となる。

### 2-2 Liu 推定量

次はデータの高次元化に伴うデータの従属性に関する検討である。データの従属性とは、いわゆる多重共線性の問題である。多重共線性は高次元データの場合にしばしば起こる問題であって、データの持つ次元がデータ変量の数より少ない場合に発生しうる。もしくは、似たような性質を持った変数がデータの中に含まれ、これらの変数のどちらが結果を表す応答変数への影響なのかを判断することができないことがある。つまり、この場合、データ内の変数間で線形的な従属関係が存在していることを意味している。これらのデータセットに対して、そのまま回帰モデルなどにデータを投入して通常の最小二乗法などで計算すると、計算不能もしくは解が不安定になる。簡単にこの現象を説明すると、データの表す行列から構成される行列の逆行列の計算において、データ行列が線形従属であると計算ができないことに起因する。このような問題が発生しないようなデータに関しては、線形モデルでは通常の最小二乗法は有用的な方法であるが，病的なデータセットに対しては Ridge 回帰，Liu 推定，Lasso などの様々なデータの従属性に対処するための方法が存在する。線形従属性の問題が 2－1 節で挙げたロジスティック回帰などの統計モデルに関しても発生しうる問題であり、これらに対応する Ridge, Liu, Lasso などの手法がロジスティック回帰モデルを含む一般化線形モデルにも問題を拡大し、一般化して適用可能であることに注意する。結果として、本検討はまず線形モデルに限った検討となるが、ロジスティック回帰に対して直接的な応用が可能であることにも注意する。本検討で着目する Liu 推定量 [17]は、Ridge 回帰に比べて、Liu 推定法が持つパラメータの変化に対し Liu 推定量が線形的に変化し、扱いやすい推定量であることが取り上げる理由に挙げられる。その Liu 推定量に対して線形制約条件が仮定された問題に対する推定法が Kaçiranlar 等[18]によって提案されている。ここでの線形制約条件とは、例えば、情報通信サービスで考えると、あるオプション A に比べ、オプション B を与えたときのほうがユーザの効用（満足度みたいなもの）に対して 2 倍の効果があることが事前にわかっている状況において、その効果を知りたいような問題を考えることがしばしば存在する。しかし、Kaçiranlar 等によって

提案された手法は、多重共線性と制約条件の下での解法という二つの問題に対して、独立に問題を捉えたアドホック的な手法であり、それを改善することが本検討の目的である。

## 3 主結果

2-1 節、2-2 節のそれぞれに対する課題とその検討結果を、それぞれ 3-1 節、3-2 節で記述する。

### 3-1 モデル評価基準

ここでは、2-1 節で取り上げたモデル評価基準についての結果を紹介する。特に、今回着目したのは、線形モデルで用いられている決定係数の一般化という観点から、一般化線形モデルに適用可能な絶対基準のモデル評価尺度に注目している。1 節に挙げたように、最終的な適用モデルは、ロジスティック回帰もしくはロジットモデルではあるが、そのモデル評価基準の統計的性質探る目的のため、下記（１）において、一般化線形モデルの一種であるポアソン回帰モデルへの適用を考える。直接的なアプローチではないが、Zheng and Agresti による RCC および Eshima and Tabata によって提案された ECC, ECD の統計的な性質は十分検討されておらず、一般化線形モデルの中でも扱いやすいポアソン回帰モデルを用いて、その統計的性質を十分に検討し、そして、その検討結果を踏まえて、ロジスティック回帰に応用した結果を（２）で説明する。また、（２）では具体的にロジスティック回帰モデルによるスパムメール判定を応用事例として紹介し、その有用性について議論する。

#### （１）ポアソン回帰モデル

ここでは、Eshima and Tabata で提案されていた ECC, ECD のポアソン回帰モデルに関する統計的性質を検討する。ECC, ECD は情報理論で用いられるエントロピーを利用したモデル評価尺度である。その二つに共通する基礎的情報量が mpp と呼ばれる値である。この mpp 自体もモデルを評価するうえでの情報が詰まっており、この mpp の性質について 3 点の観点で検討する。

１． mpp の明示的な表現
２． 単調増加性
３． 推定量の性質

得られた一部の結果を簡潔に紹介する。そもそも mpp とは対数オッズ比によって定義され，応答変数が持つ情報量が説明変量をあたえることで、どれだけその情報量が平均的に減少したのかを測る指標として定義されている。対数オッズ比で定義された値が情報理論で使われる Kullback-Leibler を用いて表現されることが過去の研究によって示されている。更には、mpp は一般化線形モデルにおける正準パラメータと応答変数の相関係数によっても定義されていることが示されている。そこで本研究では、ポアソン回帰モデルの仮定を相関係数によって定義された式に適用し、具体的に mpp の値を計算している。実際、この mpp の明示的な表現は説明変数の積率母関数とその微分によって計算されることが示される。これが mpp の明示的な表現である。次の観点は単調増加性に関するものである。一般的に、説明変量の数が増えれば、モデルとしての説明力が上がるはずである。情報を多く使うため、その説明力が上がることは合理的である。また係数パラメータの絶対値が増加すれば、一般的には、その係数に対応する説明変量が応答変数に対して影響を強く及ぼすことが想定される。そこで本検討では、mpp の係数パラメータの関数としての単調性に関する検討を、一次元の考察を経て多次元への考察へと拡張し、mpp が単調増加であるための必要十分条件を理論的に導出した。最後に推定量の性質である。以上の話は、mpp の母数(population value)の話である。実際には、決定係数のように、具体的に mpp の値を計算（推定）する必要がある。本検討では上記１で理論的に導いた mpp の明示的な表現を用いて、mpp の値を推定するための新たな推定量を提案する。具体的には、明示的に表された mpp の理論式内に現れる係数パラメータにその最尤推定量を代入した値として提案する。既存の推定量としては、標本相関係数や（不偏）標本相関係数、Jack-knife 推定量や様々な推定量が存在し得るが、これらの推定量と、本検討で用いている推定量を比べ、推定量の良さを比較する。実際、複数の既存の推定量の間に、同値関係が存在し、それを理論的に証明することができる。すべての推定量間で綺麗な同値関係が存在するわけではないため、不偏標本相関係数と本検討で提案する推定量をシミュレーションで比較する。ポアソン回帰モデルの推定に使ったサンプルの大きさは 500 であり、シミュレーション結果の図 1 の実線は既存の不偏標本相関推定量を指し、破線が提案した推定量になる。このシミュレーションは説明変量にパラメータ $\lambda=2$ の指数分布を適用した時の結果である。図 1，図 2 はそれぞれ（推定）Bias および（推定）平均二乗誤差の平方根(RMSE)を表したもので、推定 Bias，推定 RMSE に利用した Monte Carlo サンプル数は 2000 で

あり、定数項パラメータ $\alpha$ は 0.1 で固定し、各係数パラメータ $\beta$ の値に応じて比較をおこなっている。 図 1 は値が 0 に近いほうが良い推定量を表し、図 2 は絶対値が小さいほうがよい推定量を指す。どちらの指標においても、提案推定量の有効性を見ることができる。

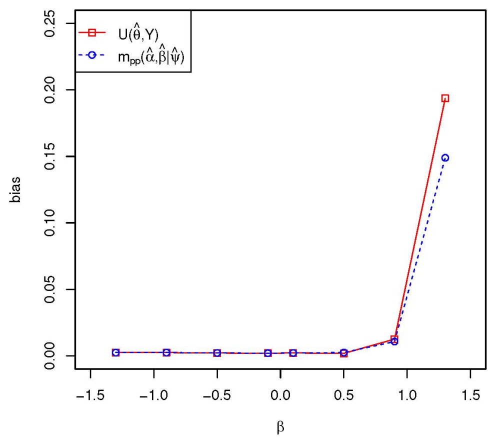


図 1：推定 Bias の比較（X～Exp(2)）

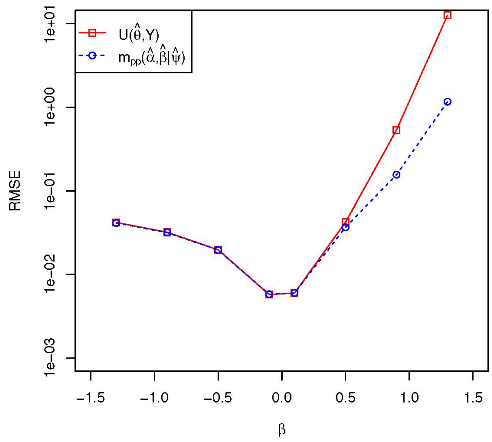


図 2：推定 RMSE の比較（X～Exp(2)）

### （２）スパムメールの判定への応用

本節では、Zheng and Agresti によって得られた RCC と呼ばれる値に対するロジスティック回帰モデルへの応用について検討する。まず、ロジスティック回帰モデルの検討に入る前に、実用例として紹介するスパムメールの判定における状況を説明する。想定している状況はスパムメールかスパムメールではないかのベルヌーイ変数の応答変数 Y（Y=1:スパム、Y=0：スパムではない）に対して、説明変数 X は、ある特定のキーワードがメール本文中に含まれているかどうかを表すベルヌーイ型の変数である。説明変量に用いるキーワードは複数存在するため、m 変量ベルヌーイ変数 **X** を仮定することになる（$m\geq 1$）。また各キーワード間にも従属性が存在する可能性があるので、パラメータ$\mathbf{p}$の次元は$2^m$ のベルヌーイ変数となる。応答変数がベルヌーイ変数となることから、仮定するモデルはロジスティック回帰モデルとする。このとき、Zheng and Agresti によって提案された RCC と呼ばれる値の推定量を計算することになる。ポアソン回帰の検討時に用いた同様の考えに基づいた推定量をロジスティック回帰モデルに適用することができる。具体的には、ロジスティックモデルの仮定の下、RCC の真の値を計算し、線形予測$\alpha+\boldsymbol{\beta}^T X$に含まれる係数パラメータ$\alpha$と $\boldsymbol{\beta}$ にそれぞれ最尤推定量を代入し，更にパラメータ$\mathbf{p}$にも最尤推定量を代入した物である。このとき、この推定量の値が

$$\sqrt{\frac{\overline{\hat{Y}^2}-\overline{Y}^2}{\overline{Y}(1-\overline{Y})}}$$

で表されることを示した。この値はm変量ベルヌーイ変数のパラメータ$\mathbf{p}$の最尤推定量を用いているのにもかかわらず、上記の値は、応答変数 Y の観測ベクトル **Y** とその予測ベクトル値$\hat{\mathbf{Y}}$のサンプル平均のみで表される簡易な統計量であることに注意する。

実際にこの計算されたモデル評価尺度を用いた実例を紹介する。今回、実験に用いたのは、UCI Repository にある Spambase Data Set [19]である。スパムの有無を表す変数とその他 57 種の説明変数からなるデータである。57 種類の説明変数の内、54 種類の説明変数を用いる。この 54 種類の説明変数のオリジナルデータはある特定の Keyword がメール本文中に何回現れるかという計数データであったが、54 種類の文字の含有の有無を表すベルヌーイ変数に変換して今回のデータ実験に用いた。想定できるモデルとしては、2＾54 という天文学的な数のモデルが考えられる。そのため、本実験データでは例示のため、各説明量を 1 つのみ用いたモデルである全 54 種類のモデルを比較対象とする。推定量の値を例示しても、その妥当性を判断しにくいため、AIC の値も同時に例示する。2 節でも説明したように AIC は相対尺度であるため、その値自身には意味を持たないが、モデル選択においては意味のあるスコアである。そのため、本検討で計算した RCC の推定値

とAICで求めた値の54種類のモデル内のランキングを例示する。しかし、図3では、すべての54種類の内10種類のモデルの値のみを示している。

| | No. 1 | No. 2 | No. 3 | No. 4 | No. 5 | No. 6 | No. 7 | No. 8 | No. 9 | No. 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| RCC | 0.239 | 0.304 | 0.335 | 0.091 | 0.408 | 0.310 | 0.522 | 0.341 | 0.298 | 0.310 |
| AIC | 5915 | 5754 | 5653 | 6136 | 5401 | 5738 | 4814 | 5642 | 5771 | 5737 |
| RCC rank | 27 | 20 | 15 | 49 | 8 | 19 | 3 | 14 | 21 | 18 |
| AIC rank | 30 | 20 | 16 | 49 | 10 | 19 | 3 | 15 | 21 | 18 |

図3：スパムデータに対するAICとRCCの値およびモデル内ランキング

結果は部分的なものであるが、RCC と AIC の全 54 種類内のモデルランキングは概ね同様のランキングである。ここに掲載していない他44種類に関しても同様の傾向である。つまり、この結果からRCCはAICと同様の働きをしつつ、決定係数と同様にその値自身にも意味を持つ統計量として計算可能であることが示される。

### 3-2 制約条件下のLiu推定量

続いて、高次元データに対応する推定量についての結果を述べる。Liu推定量は、もともとRidge回帰に対比して提案された推定量である。Ridge回帰の推定量はRidgeパラメータを持ち，そのRidgeパラメータはRidge推定量の分母に該当する部分に現れる。一方でLiu推定量は、LiuパラメータがLiu推定量の分子に該当する部分に表れる推定量となっている。かみ砕いて例示すると、関数 1／xを思い浮かべたとき、xの変化に対して、関数 1／xの値は非線形的に変化するが、一方で関数xの値はxの変化に対して線形的に振る舞うため、パラメータを変化させたときの挙動がつかみやすい。これが、Liu推定量がRidge推定量に比べて扱いやすい推定量である所以である。更に、Liu推定量の望ましい性質に、ある特定の範囲において、Ridge推定量と同様の値を持つLiu推定量がLiuパラメータを調整することで一意に定めることができるという特徴がある。つまり、Liu推定量はRidge推定量と同様の値を与えるように調整可能であり、扱いやすい推定量ということである。Liu推定量に関する研究は様々存在し、近年も盛んに検討されている。その1つにKaçiranlar等[18]による線形制約条件下 $R\boldsymbol{\beta}=\mathbf{r}$ におけるLiu推定量の提案がなされている。ただし、Rと $\mathbf{r}$ は既知の値であり、$\boldsymbol{\beta}$ が推定するべきパラメータである。しかし、Kaçiranlar等の推定量は，Liu推定量の中に現れる最小二乗推定量に単に制約条件下で得られる推定量を代入した単純なものである。Liu の考え方と制約条件下の解を使っているという点では、両方の問題を加味した推定量と考えることもできるが、代入して得られた解が尤もらしいという合理的な説明を与えることができない。一方、Kurto ˘glu等[20]はLiu 推定量がある損失関数の解で与えられることに言及した。そのため、そのアイデアを用いて、本検討では、損失関数に対して制約条件 $R\boldsymbol{\beta}=\mathbf{r}$ を課した制約条件付きの最適化問題として問題を定式化し、Lagrangeの未定乗数法を利用して、制約条件下におけるLiu推定量の新しい推定量を提案する。具体的には

$$\widehat{\beta^*_{RL}} = \widehat{\beta_L} + (X^\top X + I)^{-1}R^\top(R(X^\top X + I)^{-1}R^\top)^{-1}(r - R\widehat{\beta_L})$$

で与える推定量を考える。ただし、$\widehat{\beta_L}$ はLiu推定量で、Iは単位行列、Xはデータ行列である。本検討では、この提案する推定量とKaçiranlar等によって与えられた推定量をscalar mseの観点で理論的に比較をおこなう。具体的にはそれぞれの推定量の理論Biasと理論分散を計算して、提案する推定量がscalar mseの観点でKaçiranlar等の推定量を優越するための必要十分条件を与えた。実際、必要十分条件であるため、必ず提案の推定量が優れているという結果ではなかった。しかし、様々なシミュレーション実験を試した結果、提案する推定量が優越するケースが殆どであり、Kaçiranlar等の推定量が提案の推定量を凌駕するケースはまれであった。具体的なシミュレーション結果は図4，5に示す。共に推定scalar mseを示しており、異なる条件の下でおこなった実験結果である。共に $\boldsymbol{\beta}$ の次元は4とし，説明変量は定数項を含む4次元ベクトルでサンプルの大きさは1000とした。ただし、多重共線性を引き起こしやすい状況を作るため、定数項ではない2つの説明変量X1とX2は独立な標準正規分布から発生し、残りの定数項ではない説明変量X3はX1+X2+uとし、uはCase 1では標準正規分布、Case 2では分散0．01、平均0の正規分布から発生させた。また制約Rは共通の値を与え、$\beta$ の真の値はCase 1とCase 2で異なる値を設定している。Kaçiranlar等の推定量は黒で示し、提案の推定量は赤で示している。横軸のdはLiuパラメータの値である。また、推定scalar mseを計算するためのMonte Carloサンプル数は10000とした。推定scalar mseを比べているため、低い値の方が良い推定量を与えていることを示し、各dに対しても提案の推定量が低い推定 scalar mse を与えていることが見て取れる。また、各dに関する最適な解も検討することができ、最小のdを与える推定量で比べ

たときも提案の推定量が良い推定を与えることが見て取れる。

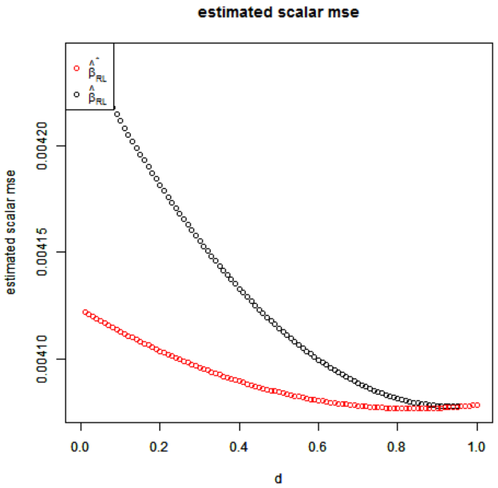

図 4：推定 scalar mse の比較(case 1)

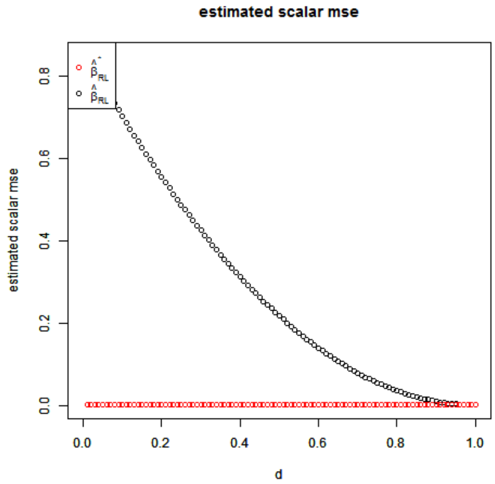

図 5：推定 scalar mse の比較(case 2)

## 3 まとめ

本研究は、特にブロードバンド・エコノミクスで利用されているロジスティック回帰やロジットモデルの統計モデルに関する課題を扱った。しかし、ここで検討した 2 つの課題は新しい課題であるため、既存の研究成果が十分ではないという現状から、ロジスティック回帰やロジットモデルに対して直接課題を解決するには少し困難な状況であった。そのため、1 つ目の課題であるモデル評価尺度に関しては、ポアソン回帰モデルにおける検討を慎重におこなった結果を十分に用いて、ロジスティック回帰モデルへの応用を行った。実際の応用例では、スパムメールの判定への適用事例を紹介したが、説明変数が多変量ベルヌーイという仮定を設定していた。この応用事例においては、多変量ベルヌーイという前提条件は特に問題のある設定ではないが、一般の分布に対しては、ロジスティック回帰モデルおよびロジットモデルに関しては、RCC や mpp といった値が、直接計算するのが困難である問題も露呈した。しかしながら、従来の標本相関等の推定量は依然として計算可能なため、実際の応用事例の一つであるスパムメールの判定に適用できるように、限定的な状況であるが多変量ベルヌーイを伴うロジスティック回帰に対するモデル評価基準を導出できたことは有用である。また、統計量の誤差を表す漸近分布の導出など様々な新たなチャレンジングな課題やポアソン回帰モデルで検討したすべての検討事項に対してロジスティック回帰モデルへの適用は完了しておらず、引き続き多数の検討事項が存在する。また、Liu 推定量に関する検討も単純な回帰モデルからの検討をおこなった。原稿中で言及したように、ロジスティック回帰モデルへの応用は拡張可能であり、本検討で得た検討をロジスティック回帰やロジットモデルを含むような一般線形モデルへの拡張をおこなっていきたい。

## 【参考文献】

[1]依田高典, ``ブロードバンド・エコノミクス―情報通信産業の新しい競争政策,'' 日本経済新聞社, 2007

[2] T. Kurosawa, A. Inoue, K. Nishimatsu, M. Ben-Akiva, and D. Bolduc, ``Customer-choice behavior modeling with latent perceptual variables,'' Intelligent Engineering Systems Through Artificial Neural Networks, vol.15, pp.419-426, 2005.

[3] T. Kurosawa, A. Inoue, and K. Nishimatsu, ``Telephone service choice-behavior modeling using menu choice data under competitive conditions,'' Intelligent Engineering Systems Through Artificial Neural Networks, vol.16, pp.711-720, 2006.

[4] T. Kurosawa, A. Inoue, and K. Nishimatsu, ``Service-choice behaviour modelling with latent perceptual variables,″ International Journal of Electronic Customer Relationship Management, vol.2, no.3, pp.228-250, 2008.

[5] T. Kurosawa, K. Nishimatsu, M. Iwashita, A. Inoue, M. Ben-Akiva, and D. Bolduc, ``Service-choice behavior modeling constructed with multiple decision-making processes,″ International Journal of Services, Economics and Management, vol.3, no.3,pp.301-322, 2011.

[6] T. Kurosawa, D. Bolduc, M. Ben-Akiva, A. Inoue, K. Nishimatsu, and M. Iwashita, ``Demand analysis by modeling choice of Internet access and IP telephony,″ International Journal of Information Systems in the Service Sector, vol.3, no.3, pp.1-26, 2011.

[7] M. Iwashita, A. Inoue, T. Kurosawa, and K. Nishimatsu, ``Microarea selection method for broadband infrastructure installation based on service diffusion process,″ International Journal on Advances in Telecommunications, vol.10, no.1/2, pp.60-71, 2017.

[8] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716-723.

[9] McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. *Econometrics*, pages 105-142.

[10] Cox, D. R. and Snell, E. J. (1989). *The Analysis of Binary Data*. Chapman & Hall, London, second edition.

[11] Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika*,439 78(3):691-692.

[12] McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman & Hall.

[13] Zheng, B. and Agresti, A. (2000). Summarizing the predictive power of a generalized linear model. *Statistics in Medicine*, 19:1771-1781.

[14] Eshima, N. and Tabata, M. (2007). Entropy correlation coefficient for measuring predictive power of generalized linear models. *Statistics & Probability Letters*, 77:588-593.

[15] Eshima, N. and Tabata, M. (2010). Entropy coefficient of determination for generalized linear models. *Computational Statistics and Data Analysis*, 54:1381-1389.

[16] Eshima, N. and Tabata,M. (2011). Three predictive power measures for generalized linear models: The entropy coefficient of determination, the entropy correlation coefficient and the regression correlation coefficient. *Computational Statistics and Data Analysis*, 55:3049-3058.

[17] K. Liu, *A new class of biased estimate in linear regression*, Communications in Statistics-Theory and Methods 22 (1993), no. 2, 393-402.

[18] S. Kaçiranlar, S. Sakallioğlu, F.Akdeniz, G.P.H.Styan, and H.J.Werner, *A new biased estimator in linear regression and a detailed analysis of the widely analysed dataset on Portland cement*, The Indian Journal of Statistics 61 (1999), no. 3, 443-459.

[19] Dheeru Dua and Efi Karra Taniskidou. UCI machine learning repository, 2017. http://archive.ics.uci.edu/ml.

[20] F. Kurto ˘glu and M.R. Özkale, *Liu estimation in generalized linear models: application on gamma distributed response variable*, Statistical Papers 57 (2016), no. 4, 911-928.

## 〈発　表　資　料〉

| 題　　名 | 掲載誌・学会名等 | 発表年月 |
|---|---|---|
| Entropy correlation coefficient for the Poisson regression model | Three-day ATMS workshop Multi- and high-dimensional statistics - Copulas - Survival analysis - Model selection | 2018 年 8 月 |
| Predictive power measures for the Poisson regression model | Fellows Research Meetings (Goulburn 35) | 2018 年 11 月 |
| 制約条件下の新 Liu 型推定量 | 第31 回日本計算機統計学会シンポジウム | 2018 年 11 月 |
| ロジスティック回帰モデルに対する回帰相関係数とスパムメール判定への応用 | 日本オペレーションズ・リサーチ学会 | 2019 年 3 月 |
| Estimators of Goodness-of-fit Measures for a Poisson Regression Model | 62nd ISI World Statistics Congress | 2019 年 8 月（予定） |